Temporal Segmentation Tool for High-Quality Real-Time Video Editing Software

Carlos Cuevas and Narciso García

Abstract — *The increasing use of video editing software requires faster and more efficient editing tools. As a first step, these tools perform a temporal segmentation in shots that allows a later building of indexes describing the video content.*

Here, we propose a novel real-time high-quality shot detection strategy, suitable for the last generation of video editing software requiring both low computational cost and high quality results. While abrupt transitions are detected through a very fast pixel-based analysis, gradual transitions are obtained from an efficient edge-based analysis. Both analyses are reinforced with a motion analysis that helps to detect and discard false detections. This motion analysis is carried out exclusively over a reduced set of candidate transitions, thus maintaining the computational requirements demanded by new applications to fulfill user needs[1].

Index Terms — Video editing software, real-time high-quality temporal segmentation, abrupt transition detection, gradual transition detection, motion analysis.

I. INTRODUCTION

The amount of consumer electronic devices including video cameras (i. e. digital cameras, mobile phones, etc.) has augmented rapidly over recent years [1], thus increasing the number of consumers who generate their own video collections. Moreover, thanks to improved computers and Internet growth, applications that work with large collections of video data have recently appeared [2] [3] (digital libraries, distance learning, video-on-demand, digital video broadcasting, interactive TV, multimedia information systems, etc.), which are commonly used by millions of users (professional and non-professional) [4]. Therefore, to meet user needs, several video editing tools for automatic indexing and retrieval of relevant material have been developed along the last years [5], which require an initial temporal segmentation into shots to proceed further on [6]. However, the temporal segmentation strategies proposed until now are not able to satisfy application demands: those that provide high quality results are computationally inefficient, while those that are fast enough do not offer satisfactory results.

Video shot transitions can be classified into two categories [7]: abrupt transitions and gradual transitions. Abrupt transitions, also known as cuts, take place between two consecutive images and are the most common transitions. On the other hand, gradual transitions are spread along a larger number of images and can further be classified into dissolve, fade-in, fade-out, and wipe [8]. While dissolves are the most common gradual transitions and fades can be treated as a special case of a dissolve, wipes are not commonly used [9]. Therefore, dissolve detection draws more attention of the researchers.

Here, we propose a novel and very efficient shot detection strategy that is able to provide high-quality real-time results. In a first stage, a pixel-based gap analysis is applied, allowing the detection of abrupt transitions between shots. In parallel, the amount of significant edge points along the images is analyzed, allowing the location of gradual transitions. Both analyses are reinforced with a second stage based on motion analysis, which is applied only on the previously detected transitions. This second stage allows to reduce the problem of threshold selection and, furthermore, it does not increase significantly the overall computational cost since the analysis of motion, which a priori may seem computationally expensive, applies only to some preselected images. Therefore, this proposal is more suitable than previous shot detection strategies for applications embedded in the last generation of consumer electronic devices.

To evaluate the quality and computational efficiency of the proposed strategy, it has been tested over a database of more than 30 sequences. These sequences have a duration of approximately 3 hours, contain more than 1000 shots, and have certain characteristics that hinder the proper behavior of classical shot detection algorithms such as large moving objects in the scene or quick and varied camera motion.

The rest of the paper is organized as follows. Section II contains a brief state of the art concerning the shot transition detection strategies. Section III details the proposed system architecture. Section IV and Section V describe the strategies for detecting, respectively, abrupt and gradual transitions. The description of the motion analysis performed on pre-detected transitions is detailed in Section VI. Finally, Section VII and Section VIII contain the obtained results and the conclusion, respectively.

II. Related Work

Numerous efficient shot boundary detection algorithms have been proposed and validated during the last decade [10]. Among these, some recent studies that analyze and compare the most relevant strategies can be found [11] [12] [13]. First proposals were focused on the detection of abrupt transitions (cuts) but, as these were located more efficiently, posterior strategies began to consider the detection of gradual transitions (dissolves, fades and wipes), which identification is more complex and difficult due to the many existing types of gradual transitions [14].

The early works [15] [16] proposed the computation of the differences between pixels at the same coordinates from consecutive images and determined the presence of a transition when the differences exceed the value of predefined thresholds. These strategies, despite being simple, do not provide good enough results, leading to a high number of false detections and ignoring many real transitions [17].

More recently histogram-based techniques were proposed [8] to try and reduce the number of false detections due to the existence of global movements. These strategies perform a statistical analysis of different pixel characteristics along the images and use several predefined thresholds to determine the presence of a transition [18].

As an alternative to the pixel-based and histogram-based techniques, some strategies incorporating more complex statistical analyses and using different color set spaces have been proposed along the recent years [14] [19]. These proposals improve the results of the abovementioned methods, but only working on the same kind of video sequences (news, sports, etc.) [20].

As false detections are mainly due to camera movement or moving objects in the scene, numerous motion-based strategies have been also proposed [12]. Some of these strategies apply motion compensation techniques before carrying out the analysis of the pixels [21], while others analyze the displacement of singular points extracted from the images [22].

Finally, another proposed alternative is to make use of different techniques simultaneously [23]. In this way, by taking the advantages of each one of the combined techniques, high quality results are obtained [24].

On the one hand, pixel-based methods, histogram-based methods, and those based on more complex statistical analysis, are quick and easy to implement but have some drawbacks that must be taken into account [25], like selecting the best threshold according to the characteristics of the analyzed video. If the amount of motion in the video is important, high thresholds should be applied whereas, if there is no significant motion, thresholds should be low [13]. Therefore, it is necessary to select different threshold values depending on the characteristics of each sequence and even for different parts of the same sequence. If the chosen value is too low the amount of false detections will be higher. Conversely, if the threshold is too high a large amount of transitions will be overlooked [12].

On the other hand, methods based on motion analysis and those that combine several strategies are able to get better results,

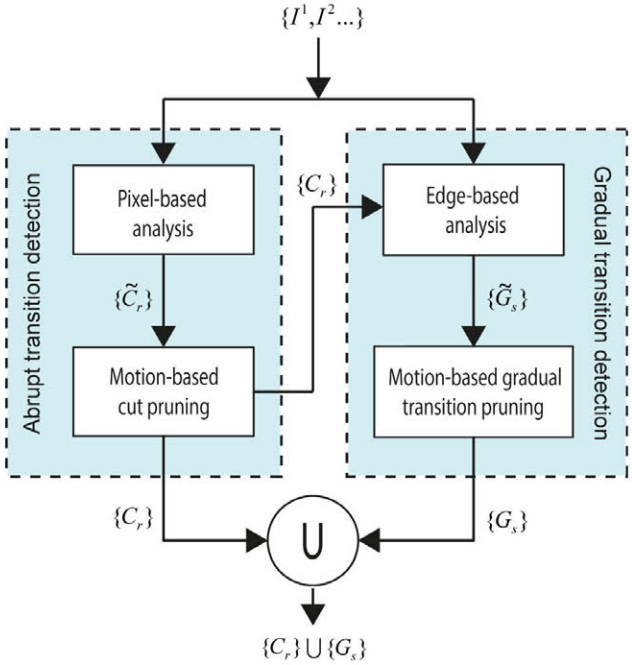

Fig. 1. Block diagram for abrupt and gradual transition detection.

detecting most of the existing transitions and reducing the amount of false detections [26] . However, these methods are much more complex and considerably slower than previous ones [27]. In addition, these strategies also require thresholds. So, the problem of selecting appropriate values for these thresholds is still present.

III. System Overview

The proposed system combines two strategies that use low level techniques with a motion analysis between pairs of preselected images. Fig. 1 shows a block diagram containing a detailed description of the processing modules in that system. The input to the system is the image sequence under analysis, $\{I^1, I^2 \ldots\}$, which is introduced into two parallel processing lines. The loop at the left of the figure contains the stages used for the detection of abrupt transitions, while the one at the right shows the stages for the detection of gradual transitions.

To locate the existing abrupt transitions, we begin analyzing the intensity differences between consecutive images at the pixel level. As a result of this analysis, a set of candidate abrupt transitions, $\{\tilde{C}_r(I^{r1}, I^{r2})\}$, is obtained, where r is the index of the candidate transitions and (I^{r1}, I^{r2}), with $r_2=r_1+1$, are the images delimiting the r-th transition. In a second step, we apply a motion analysis on each pair of candidate images to decide whether each candidate transition really exists or is a false detection. If a candidate transition is classified as correct it is added to the set of final abrupt transitions, $\{C_r(I^{r1}, I^{r2})\}$.

In the case of gradual transitions, the first step is to extract gradient information from each image in the sequence. The analysis of this information starts whenever an abrupt transition is identified and it takes place between the last image whose gradients were previously analyzed and the newly detected abrupt transition. From this analysis we obtain pairs of images, (I^{s1}, I^{s2}), that delimit candidate gradual transitions, $\{\widetilde{G}_s(I^{s1}, I^{s2})\}$, where s is the index of each candidate gradual transition. To identify and separate false detections from correct ones we apply a motion analysis stage, similar to that used in the detection of abrupt transitions, obtaining the final set of gradual transitions, $\{G_s(I^{s1}, I^{s2})\}$. The result of the system is the union of both sets of final transitions, $\{C_r\} \bigcup \{G_s\}$.

It is necessary to consider that although the proposed system is able to work at high speed, there is a latency period at the beginning of the gradient analysis stage (gradient analysis starts only when a new abrupt transition is detected). Nevertheless, as this latency depends on the maximum difference between consecutive abrupt transitions, it is possible to ensure a maximum latency by inserting artificial abrupt transitions.

IV. Abrupt Transition Detection

The first step in the proposed strategy is a fast and efficient pixel-based algorithm, which allows to locate the candidate abrupt transitions along the sequences. Usual pixel-based algorithms stem from the idea that consecutive images belonging to the same shot are more similar than consecutive images belonging to different shots [12]. However, the evaluation of differences between consecutive images has some problems, such as the appearance of false detections due to the presence of illumination changes. In these situations a large amount of pixels suffer significant intensity variations between consecutive images, which leads to the detection of false transitions.

To solve this problem, we propose the use of a novel and powerful metric that is invariant to illumination changes, as it compares the intensity variations between pairs of consecutive images, (I^n, I^{n-1}), with respect to their mean intensities, (μ^n, μ^{n-1}). This metric is defined as

$$M_p(I^n) = \frac{1}{HW} \sum_{h,w} \rho_{h,w}^n , \quad (1)$$

where H and W are the height and the width of the compared images, (h,w) are the spatial coordinates of each pixel, and ρ is defined as

$$\rho_{h,w}^k = \begin{cases} 1, & \text{if } \operatorname{sign}(\Delta I_{h,w}^n) = \operatorname{sign}(\Delta I_{h,w}^{n-1}) \wedge |\Delta I_{h,w}^n| > T_{n1} \\ -1, & \text{if } \operatorname{sign}(\Delta I_{h,w}^n) \neq \operatorname{sign}(\Delta I_{h,w}^{n-1}) \wedge |\Delta I_{h,w}^n| > T_{n1} \\ 0, & \text{otherwise} \end{cases} , \quad (2)$$

where ∧ represents a logical AND, $\Delta I_{h,w}^n = I_{h,w}^n - \mu^n$, and T_{nl} is a noise threshold, with a typical value lower than 3, which avoids taking into account small intensity variations.

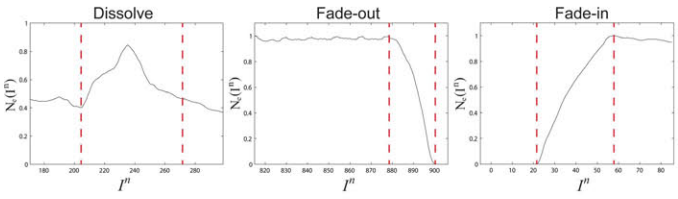

Fig. 2. Amount of edge-points along different types of gradual transitions.

The overall algorithm for abrupt transition detection can be given as follows:

- For each new input image, I^n, apply the metric described in equation (1).
- Compare the values of $M_p(I^n)$ between I^{n-N_w} and I^n, and locate the local minimum, $m_{M_p} = M_p(I^\theta)$, where θ is the position of the minimum and N_w determines the number of images being compared.
- A new candidate abrupt transition, defined by $I^{r1} = I^{\theta-1}$ and $I^{r2} = I^\theta$, is added to the set $\{\widetilde{C}_r\}$ if

$$\left| m_{M_p} - \min\{M_p(I^i) / i \in [n - N_w, n] \wedge i \neq \theta\} \right| > T_p , \quad (3)$$

where ∧ represents a logical AND, and T_p is the threshold that determines when any of the values of $M_p(I^n)$ stands out from its environments.

V. Gradual Transition Detection

The rationale behind the proposed strategy for the detection of potential gradual transitions is to compute the evolution of image gradients (edge-points) along the video sequences.

In a dissolve-type gradual transitions the first shot gradually disappears while the second one gradually appears. On this basis, the evolution of the edges along a video sequence can give us a hint about the location of gradual transitions. Numerous experiments have demonstrated that the amount of edge-points along a gradual transition evolves from those belonging to the first shot to those in the second one. In between, edge-points from both shots appear, generating a local maximum during the transition. Then, identifying this type of variations along a video sequence we can locate its gradual transitions.

In the case of fade-ins or fade-outs similar behaviors can be observed. In a fade-in, starting from zero, the amount of edge-points increases progressively along the transition, while in a fade-out the amount of edges decreases progressively to become zero. Then, these transitions can be also detected by analyzing the evolution of the edges.

Fig. 2 depicts some examples of the evolution of the amount of edges along the three abovementioned gradual transitions: dissolve, fade-in, and fade-out. The vertical

dashed lines mark the beginning and the end of each transition.

The overall algorithm for gradual transition detection can be given as follows:

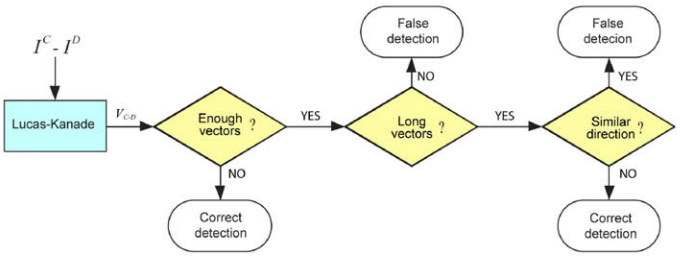

Fig. 3. Flow chart for the motion-based analysis.

- Calculate the percentage of image pixels, $N_e(I^n)$, with a gradient higher than a noise threshold, T_{n2}, between the last pair of detected abrupt transitions.
- Apply a low pass filter on $N_e(I^n)$ to remove noise variations and normalize the result of the filtering.
- Locate the set of local maxima, $\{M_{N_e}(I^n)\}$, and their anterior and posterior closest minima, $\{M_{N_e}(I^{n-A})\}$ and $\{M_{N_e}(I^{n+B})\}$.
- A new candidate gradual transition, defined by $I^{s1} = I^{n-A}$ and $I^{s2} = I^{n+B}$, is added to the set $\{\widetilde{G}_s\}$ if either of the following two conditions are fulfilled:

$$\begin{aligned} &\left|M_{N_e}(I^n) - M_{N_e}(I^{n-A})\right| > T_e \ \wedge \ A \geq \frac{N_g}{2}, \\ &\left|M_{N_e}(I^n) - M_{N_e}(I^{n+B})\right| > T_e \ \wedge \ B \geq \frac{N_g}{2} \end{aligned} \tag{4}$$

where $\wedge$ represents a logical AND, T_e is the minimum percentage variation of $N_e(I^n)$ to consider a candidate transition, and N_g determines the minimum number of images of a gradual transition to be detected.

VI. Motion-Based Pruning

Previously described strategies are able to detect most abrupt and gradual transitions in real-time. However, they do not avoid some false detections resulting from undesirable situations such as fast camera displacements, zooms, etc. To detect and separate these false detections from the correct ones, we propose an efficient and innovative motion analysis applied over the candidate transitions resulting from the pixel-based and edge-based algorithms.

Usually, motion-based analyses provide high quality detections, but they are computationally inefficient [27]. Nevertheless, the proposed motion analysis is carried out exclusively over a reduced amount of candidate transitions ($\{\widetilde{C}_r\}$ and $\{\widetilde{G}_s\}$) and, consequently, the computational requirements of the proposed strategy are maintained.

The flowchart of the proposed motion-based strategy is detailed in Fig. 3. A set of λ motion vectors, $\{v_1...v_\lambda\}$ is obtained by applying a Lucas-Kanade pyramidal algorithm [28] over each pair of images $(I^C - I^D)$ delimiting

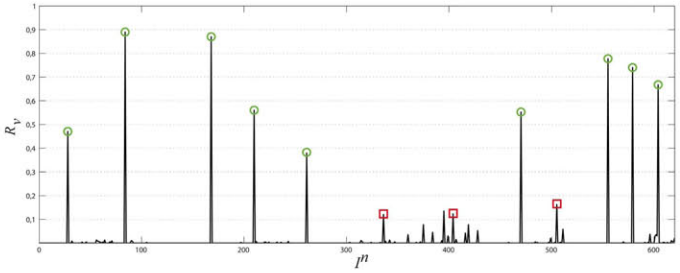

Fig. 4. Amount of motion vectors along a sequence with 615 frames. Circles mark real transitions and squares mark false detections.

each candidate transition. These vectors relate a set of M singular points obtained with Harris detector [29] in I^C, $\{s_1...s_M\}$, with points in I^D. To determine if a candidate transition is a real transition or a false detection (see the flow chart in Fig. 3), we analyze the amount, the length, and the direction of the motion vectors.

A. Amount of vectors

Firstly, the ratio R_v=1-λ/M is computed. If I^C and I^D belong to different shots, the amount of motion vectors relating points between them is much lower than the quantity of starting points. So, R_v results in large values. Nevertheless, if the analyzed images belong to the same shot, as they contain more common features, the amount of motion vectors is much closer to the number of starting singular points, thus resulting in low values of R_v. Consequently, the ratio R_v allows to identify a large amount of correct detections. The graphic in Fig. 4 shows the values of R_v along a sequence with 615 frames. The data in this graphic allow to check that images belonging to a same shot result in much lower R_v values than images from different shots.

B. Length of the vectors

In a second step, to discard false detections the mean length of the vectors, S_v, is analyzed, which is computed as

$$S_v = \frac{1}{\lambda}\sum_{i=1}^{\lambda}\left(\left(\frac{L_{H,i}}{H}\right)^2 + \left(\frac{L_{W,i}}{W}\right)^2\right)^{1/2}, \tag{5}$$

where ($L_{H,i}$, $L_{W,i}$) are the components (rows and columns) of the i-th motion vector.

As is shown in Fig. 5.a, false detections resulting from large moving objects and camera motion cause large amounts of short vectors, leading to low values of S_v. On the other hand, most of the motion vectors from images belonging to different shots are large, leading to higher values of S_v, as can be appreciated in the graphic represented in Fig. 6, which shows the values of S_v along a sequence with 650 frames. Therefore, analyzing the values of S_v, false detections resulting from

large moving objects or camera motion can be easily discarded from correct ones.

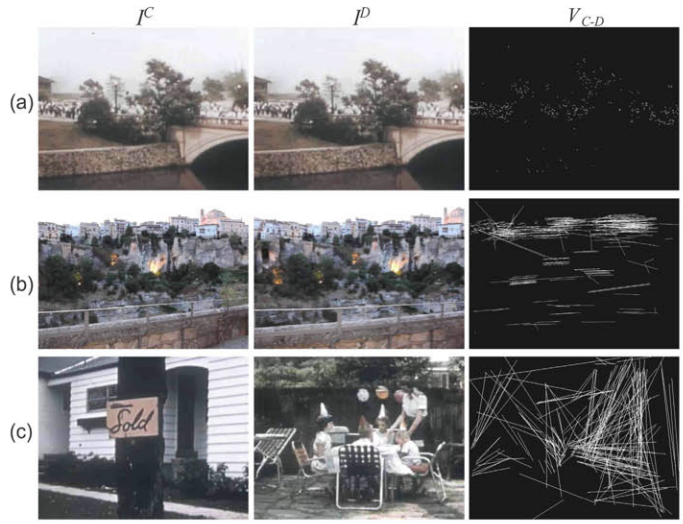

Fig. 5. Motion vectors for three candidate transitions. (a) False detection resulting from camera motion. (b) False detection resulting from a pan camera effect. (c) Correct detection.

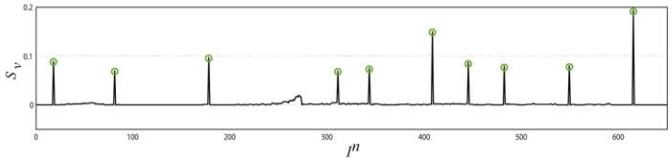

Fig. 6. Mean motion vector length along a sequence with 650 frames. Circles mark real transitions.

TABLE I
DESCRIPTION OF THE TEST SEQUENCES

Kind of sequence	Duration (sec.)	Num. of images	Abrupt trans.	Gradual trans.	Total trans.
Cartoons	904	22640	279	7	286
Musicals	1244	33112	206	8	214
Reports	1589	40629	82	77	159
Other	7012	171255	284	1	285
Test-Set	10749	267636	851	93	944

C. *Direction of the vectors*

Finally, to identify false detections resulting from fast camera changes (traveling, pans, tilts or zooms), we analyze the typical deviation of the set of vectors,

$$\sigma_v = \frac{\sum_{i=1}^{\lambda} \left(D_{v,i} - \mu_v\right)^2 \left(L_{H,i}^2 + L_{W,i}^2\right)^{1/2}}{\sum_{i=1}^{\lambda} \left(L_{H,i}^2 + L_{W,i}^2\right)^{1/2}}, \tag{6}$$

where $D_{v,i}$ is the direction of the i-th vector and μ_v is the mean direction of the set of motion vectors.

The abovementioned camera changes produce false detections where most motion vectors have similar orientations (Fig. 5.b). Consequently, these false detections result in significantly lower σ_v values than correct detections, where the direction of the obtained vectors is much more random (Fig. 5.c). Therefore, analyzing the direction of the motion vectors the candidate transitions with significant and large motion vectors can be finally classified as correct transitions or false detections.

VII. RESULTS

To test the proposed temporal video segmentation system we have analyzed a database of more than 30 sequences, which contain more than 1000 shots and a total duration of about 3 hours. Most of these sequences are cartoons, musicals and reports. These kinds of videos contain large amounts of unfavorable situations, such as: similar backgrounds in consecutive shots, camera motion, large moving objects in the scene, and illumination changes. Additionally, the frequency of abrupt and gradual transitions they contain is very high. Thus, their temporal segmentation is particularly difficult. Moreover, we have considered the union of all these sequences (labeled as *Test-Set*) to provide overall results. Table I presents a summary of the main features of the used sequences, grouped by category. The data in this table show that the amount of abrupt transitions along the described sequences is almost 10 times higher than the number of gradual transitions, since gradual transitions are far less frequent in most kind of sequences.

As evaluation measures, because they are very commonly used in the literature [26], we have used the *Recall* (R) and *Precision* (P) percentages,

$$R = 100\frac{CD}{CD + ND}\% \qquad P = 100\frac{CD}{CD + FD}\%, \tag{7}$$

where CD is the number of correct detections, ND is the amount of not detected transitions, and FD is the number of false detections. *Recall* measures the capability in detecting correct transitions, while *Precision* measures the ability in preventing false alarms. Besides these two percentages we have used a third rate, called F [30], which jointly evaluates the *Recall* and *Precision* results. This rate is defined as

$$F = 2\frac{R \times P}{R + P}\%. \tag{8}$$

A. *Abrupt Transition Detection*

In the first stage for abrupt transitions detection, the selection of appropriate values for the variables N_w and T_p is essential. For this reason an analysis of the results obtained with different values of these two variables has been carried out. The summary of the *Recall*, *Precision* and F percentages for different values of these parameters is depicted in Fig. 7.

The results of the graphs in this figure show that the number of undetected transitions increases with increasing values of N_w and T_p. Since the purpose of this stage is to detect the maximum number of existing abrupt transitions (highest *Recall*), we have decided to use N_w=5 and T_p=0.3.

Lower values of either of the two parameters do not improve the *Recall* and, nevertheless, the number of false detections increases (lower *Precision*).

The chosen values provide the highest *Recall*, but a significant amount of false detections. However, most of these false detections will be detected and discarded in the following motion-based analysis. Consequently, highest *F* rates are not obtained in this first stage but after the pruning of false detections through the motion analysis.

The final results for the application of this first stage by using the abovementioned values are summarized in Table II.

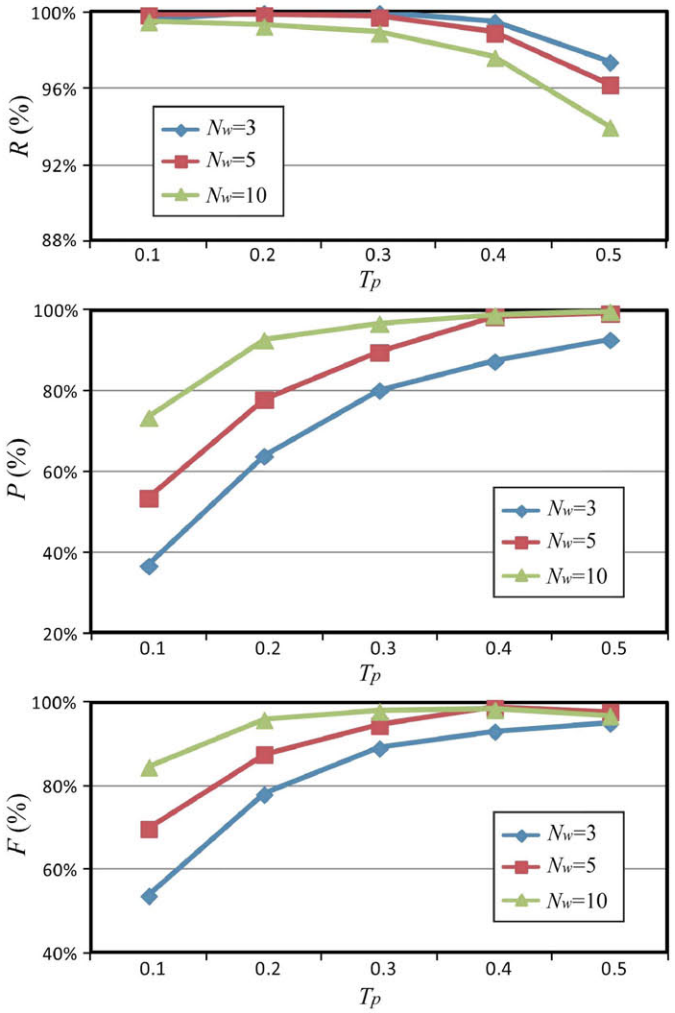

Fig. 7. ***Recall***, ***Precision*** **and** ***F*** **percentages with different values of N_w and T_p.**

TABLE II
RECALL, PRECISION, AND *F* PERCENTAGES OBTAINED WITH THE PROPOSED ABRUPT TRANSITION DETECTION STRATEGY, BEFORE MOTION ANALYSIS

Kind of sequence	*CD*	*FD*	*ND*	*R* (%)	*P* (%)	*F* (%)
Cartoons	279	50	0	100	84.80	91.78
Musicals	206	16	2	99.03	92.73	95.77
Reports	82	7	0	100	92.13	95.91
Other	284	26	0	100	91.61	95.62
Test-Set	851	99	2	99.76	89.56	94.39

B. *Gradual Transition Detection*

Similarly to the previous analysis of N_w and T_p, we have carried out in this case an analysis of the results obtained with different values of N_g and T_e. The results obtained by assigning different values to these two parameters are compared in the graphs of Fig. 8. These results show us that, using increasing values of N_g and T_e, the number of

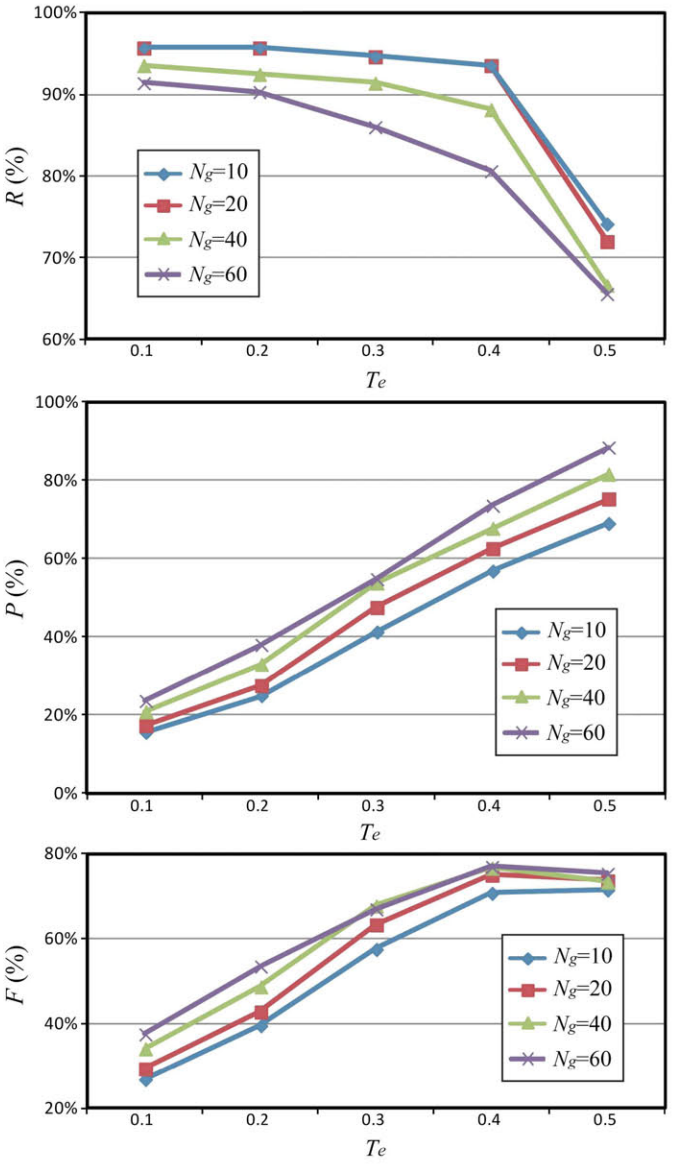

Fig. 8. ***Recall***, ***Precision*** **and** ***F*** **percentages with different values of N_g and T_e.**

TABLE III
RECALL, PRECISION, AND *F* PERCENTAGES OBTAINED WITH THE PROPOSED GRADUAL TRANSITION DETECTION STRATEGY, BEFORE MOTION ANALYSIS

Kind of sequence	*CD*	*FD*	*ND*	*R* (%)	*P* (%)	*F* (%)
Cartoons	6	11	1	85.71	35.29	50.00
Musicals	8	9	0	100	47.06	64.00
Reports	72	19	5	93.51	79.12	85.71
Other	1	13	0	100	7.14	13.33
Test-Set	87	52	6	93.55	62.59	75.00

undetected transitions increases (lower *Recall*) and the amount of false detections is reduced (higher *Precision*). As in the case of abrupt transitions detection stage, the purpose of this stage is to obtain *Recall* percentages as high as possible. Based on this purpose we have decided to use N_g=20 and T_e=0.4 since, using lower values, the *Recall* improvement is negligible and, however, the *Precision* decreases considerably. Using these values, as it is shown in the third graphic in Fig. 8, the achieved F is almost the best possible in this stage. However, for the same reasons outlined for the pixel-based analysis, F will be significantly improved after applying the motion analysis stage.

TABLE IV
RECALL, PRECISION, AND F PERCENTAGES IN ABRUPT TRANSITION DETECTION AFTER MOTION ANALYSIS

Kind of sequence	*CD*	*FD*	*ND*	*R* (%)	*P* (%)	*F* (%)
Cartoons	274	16	5	98.21	94.48	96.31
Musicals	200	4	6	97.09	98.04	97.56
Reports	76	0	6	92.68	100	96.20
Other	283	0	1	99.65	100	99.82
Test-Set	833	20	18	97.88	97.66	97.77

TABLE V
RECALL AND PRECISION PERCENTAGES IN GRADUAL TRANSITION DETECTION AFTER MOTION ANALYSIS

Kind of sequence	*CD*	*FD*	*ND*	*R* (%)	*P* (%)	*F* (%)
Cartoons	6	3	1	85.71	66.67	75
Musicals	8	3	0	100	72.73	84.21
Reports	71	3	6	92.21	95.95	94.04
Other	1	2	0	100	33.33	50
Test-Set	86	11	7	92.47	88.66	90.53

TABLE VI
FINAL GLOBAL RESULTS: DETECTION OF ABRUPT AND GRADUAL TRANSITIONS AFTER MOTION ANALYSIS

Kind of sequence	*CD*	*FD*	*ND*	*R* (%)	*P* (%)	*F* (%)
Cartoons	280	19	6	97.90	93.65	95.73
Musicals	208	7	6	97.20	96.74	96.97
Reports	147	3	12	92.45	98.00	95.15
Other	284	2	1	99.65	99.30	99.47
Test-Set	919	31	25	97.35	96.74	97.04

TABLE VII
MEAN TIMES PER IMAGE AND FINAL SPEED, CORRESPONDING TO SEQUENCES WITH DIFFERENT SPATIAL RESOLUTION

Dimensions ($H \times W$)	T_a (ms)	T_g (ms)	T_m (ms)	T_t (ms)	Speed (fps)
182 x 320	1.72	1.37	0.10	3.18	314.26
270 x 480	3.46	2.77	0.04	6.27	159.43
360 x 480	4.49	3.81	0.06	8.36	119.58
360 x 640	5.01	4.46	0.87	10.64	94.00
576 x 720	8.86	7.56	0.28	16.69	59.90

The summary of the results obtained at this stage, using the aforementioned values of N_g and T_e, is shown in Table III. The data in this table show that the proposed edge-based analysis allows the detection of most existing gradual transitions. However, this analysis also results in a high amount of false detections. Most of these false detections will be discarded in the following motion-based stage.

A. Motion-based Pruning

As can be seen in the results shown in Table II and Table III, using the pixel-based (abrupt transitions) and gradient-based (gradual transitions) proposed strategies, most shot transitions are satisfactory identified (*Recall* percentages close to 100%). However, as shown by the lower *Precision* percentages in these tables, the mentioned strategies do not prevent the detection of a high amount of false transitions. Consequently, to identify and discard most of these false detections, the motion-based strategy described in Section VI is necessary.

The results obtained after the application of this strategy appear in Table IV (in the case of abrupt transitions) and in Table V (in the case of gradual transitions). Table VI shows the final global results provided by the overall proposed strategy. On the one hand, the results in these tables show that, after the motion-based pruning, most false detections have been discarded. So, the obtained *Precision* percentages are very high in the four video categories. On the other hand, due to the erroneous detection of coherent motion between the images delimiting some candidate transitions, the number of correct detections has been slightly, but not significantly, decreased. However, as it is shown by the F rates in these tables, the global quality of the results has been highly improved.

A. Computational Efficiency

To analyze the computational efficiency of the developed system we have used a 2.66 GHz CPU with and 4 GB RAM and an optimized software implementation.

Table VII shows the mean times per image and the final speed, in terms of frames analyzed per second (fps), obtained in the analysis of various sequences with different spatial resolution. The second column of the table contains the mean time corresponding to the pixel-based analysis for abrupt transition detection (T_a). The third column presents the mean time in the edge-based analysis for gradual transition detection (T_g). The fourth column shows the mean time in the motion-based analysis (T_m). The fifth column contains the total mean-times (T_t), obtained as the sum of the previous three. At last, the sixth column presents the final achieved speeds. These results show us that the achieved speed is very high even for the sequences with larger spatial resolution. We can also observe that, on account of the fact that motion analysis is performed only on a few pairs of images per sequence, it barely increases the total computational cost, making possible the use of the proposed strategy in last generation of video editing software tools requiring real-time processing. Additionally, it should be noted that the computational costs of the pixel-based and edge-based analyses are proportional to the spatial resolution of the sequences. However, the cost in the motion analysis stage depends on the amount of candidate transitions resulting from the previous stages.

B. *Comparison with other methods*

Finally, we have analyzed the performance of the proposed strategy (in terms of running speed and quality of the results) with other alternative shot detection strategies:

- On the one hand, we have contrasted our strategy with two simple methods focused to be very fast. The first one [31] is based on the sum of absolute differences (SAD), while the second one [18] is based on the analysis of color histograms (CH).
- On the other hand, we have also compared our strategy with two algorithms that analyze the motion along the sequences: the first one [32] using salient feature matching (SFM) and the second one [21] using exhaustive block matching (EBM).

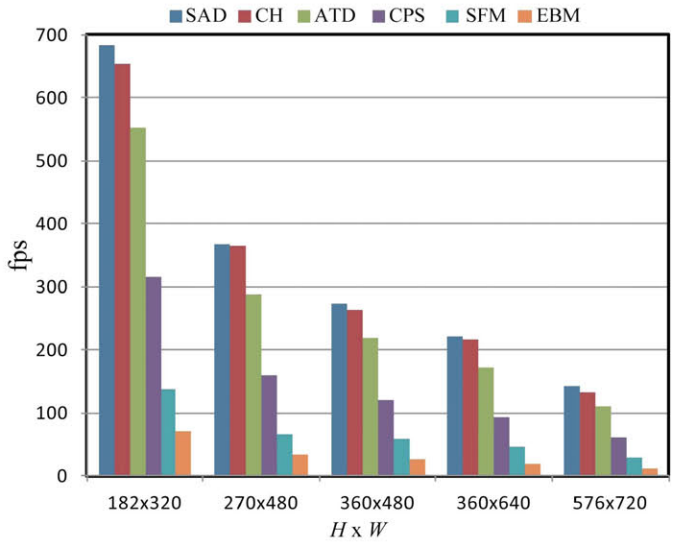

Fig. 9. Speed, in terms of frames per second, obtained with the compared shot detection strategies.

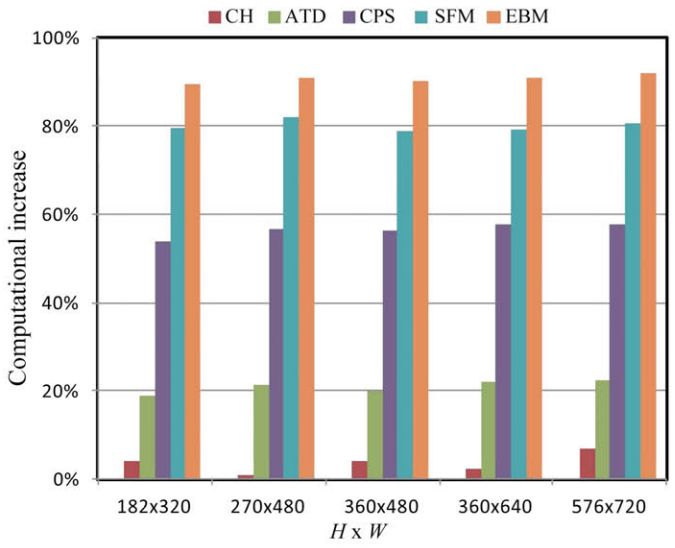

Fig. 10. Computational increase of the analyzed strategies with respect to the fastest one (SAD).

Since SAD and CH only detect abrupt transitions, we have compared them not only with the complete proposed strategy (CPS), but with a light version of it that only detects abrupt transitions (ATD).

Fig. 9 displays the values of the speed achieved with the six compared methods and Fig. 10 presents the computational increase of each method with respect to the fastest one.

The results in these figures show that the previous strategies SAD and CH are slightly faster than ATD, since they use very simple algorithms that only detect abrupt transitions. However, these previous quick strategies depend on several thresholds, which make difficult their use. The performed experiments have revealed that using thresholds to obtain *Recall* values greater than 90%, the *Precision* values are lower than 80%. So, SAD and CH are not able to achieve so good results as our pixel-based strategy to detect abrupt transitions. It should be noted that the reason why the complete proposed strategy (CPS) is considerably slower than SAD and CH is that CPS not only detects abrupt transitions but also the gradual ones.

In contrast to SAD and CH, the evaluated alternative motion-based approaches (SFM and EBM) have proven to be able to obtain almost as good results as ours: *Recall* and *Precision* percentages higher than 92% in most sequences. However, since the proposed motion analysis is applied only to some preselected images, our strategy is much faster than these previous methods. Additionally, SFM and EBM are much more complex than the proposed strategy and they also require the selection of adequate threshold values depending on the characteristics of each sequence.

In summary, the proposed strategy has shown to be the best compromise between computational efficiency, usability and quality.

VIII. Conclusion

An innovative system for the temporal segmentation of video sequences has been presented, which can efficiently detect both abrupt and gradual transitions between shots. A set of candidate abrupt transitions is obtained through a very fast analysis of the differences between consecutive images at the pixel level. In parallel, analyzing the variations of significant edge point of the images, we obtain a set of candidate gradual transitions. These two analyzes have been reinforced with a second stage based on a motion analysis, which is applied exclusively on the candidate sets of transitions. This second stage simplifies the problem of threshold selection while preserving the computational requirements of the system.

The proposed system has been tested on a wide range of test sequences, providing high quality results in unfavorable situations, such as: similar backgrounds in consecutive shots, significant changes in the content of the sequences, camera motion, large moving objects in the scene, and illumination changes. In addition, as the motion-based analysis is applied only on a few pairs of preselected images (candidate transitions), the system has proven capable to operate at very high speed.

The high-quality real-time results provided by our system demonstrate that it is more adequate than previous shot detection strategies for video editing software tools where speed and quality are required.

REFERENCES

[1] K. Sangani, "2010 gadget census [Consumer Tech Census]," *Engineering & Technology*, vol. 5, no. 14, pp. 28-29, 2010.

[2] Z. Cao and M. Zhu, "An efficient video similarity search algorithm," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 751-755, 2010.

[3] K. Seo, S. J. Park, and S. Jung, "Wipe scene-change detector based on visual rhythm spectrum," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 831-838, 2009.

[4] W. T. Peng et al., "Editing by viewing: automatic home video summarization by viewing behavior analysis," *IEEE Transactions on Multimedia*, vol. 13, no. 3, pp. 539-550, 2011.

[5] J. Han, "Object segmentation from consumer videos: a unified framework based on visual attention," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, pp. 1597-1605, 2009.

[6] M. Mehrabi, F. Zargari, and M. Ghanbari, "Fast and low complexity method for content accessing and extracting DC-pictures from H. 264 coded videos," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1801-1808, 2010.

[7] L. Chaisorn, C. Manders, and S. Rahardja, "Video retrieval – evolution of video segmentation, indexing and search," *IEEE International Conference on Computer Science and Information Technology*, pp. 16-20, 2009.

[8] O. Kucuktung, U. Gudukbay, and O. Ulusoy, "Fuzzy color histogram-based video segmentation," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 125-134, 2010.

[9] Y. Wang, Y. Yang, T. Ren, and G. Wu, "A motion-insensitive dissolve method with SURF," *IEEE International Conference on Image and Graphics*, pp. 451-456, 2009.

[10] R. Tapu and T. Zaharia, "A complete framework for temporal video segmentation," *IEEE International Conference on Consumer Electronics*, pp. 156-160, 2011.

[11] C. Snoek and M. Worring, "Multimodal video indexing: a review of the state-of-the-art," *Multimedia Tools and Applications*, vol. 25, no. 1, pp. 5-35, 2005.

[12] D. Brezeale and D. Cook, "Automatic video classification: a survey of the literature," *IEEE Transactions on Systems Man and Cybernetics - Part C*, vol. 38, no. 3, pp. 416-430, 2008.

[13] A. Smeaton, P. Over, and A. Doherty, "Video shot boundary detection: seven years of TRECVid activity," *Computer Vision and Image Understanding*, vol. 114, no. 4, pp. 411-418, 2010.

[14] V. Chasanis, A. Likas, and N. Galatsanos, "Simultaneous detection of abrupt cuts and dissolves in videos using support vector machines," *Pattern Recognition Letters*, vol. 30, no. 1, pp. 55-65, 2009.

[15] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full-video search for object appearances," *Journal of Information Processing*, vol. 15, no. 2, p. 316, 1992,

[16] H. Zhang, A. Kankanhalli, and S. Smoliar, "Automatic partitioning of full-motion video," *Multimedia Systems*, vol. 1, no. 1, pp. 10-28, 1993.

[17] S. Lawrence, D. Ziou, M. F. Auclair-Fortier, and S. Wang, "Motion-insensitive detection of cuts and gradual transitions in digital video," *Pattern Recognition and Image Analysis*, vol. 14, no. 1, pp. 109-119, 2004.

[18] Y. Xiao-juan and F. Hong-cai, "An abrupt shot change detection algorithm based on the YUV space," *International Conference on Electrical and Control Engineering*, pp. 4630-4633, 2010.

[19] M. Padalkar and M. Zaveri, "Dissolve detection based shot identification using singular value decomposition," *IEEE International Conference on Mathematical/Analytical Modelling and Computer Simulation*, pp. 312-316, 2010.

[20] B. Ionescu, V. Buzuloiu, P. Lambert, and D. Coquin, "Improved cut detection for the segmentation of animation movies," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 641-644, 2006.

[21] C. Wang, Z. Sun, and K. Jia, "Abrupt cut detection based on motion information," *IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 344-347, 2011.

[22] J. Li, Y. Ding, Y. Shi, and W. Li, "A divide-and-rule scheme for shot boundary detection based on sift," *International Journal of Digital Content Technology and its Applications*, vol. 4, no. 3, pp. 202-214, 2010.

[23] J. Sánchez and X. Binefa, "Shot segmentation using a coupled Markov chains representation of video contents," *Pattern Recognition and Image Analysis*, pp. 902-909, 2003.

[24] G. Ciocca, "A robust multi-feature cut detection algorithm for video segmentation," *Electronic Letters on Computer Vision and Image Analysis*, vol. 9, no. 1, pp. 32-46, 2010.

[25] P. Mohanta, S. Saha, and B. Chanda, "A model based shot boundary detection technique using frame transition parameters," *IEEE Transactions on Multimedia*, vol. 13, no. 6, pp. 1-11, 2011.

[26] S. Lian, "Automatic video temporal segmentation based on multiple features," *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, vol. 15, no. 3, pp. 469-482, 2011.

[27] Y. Gao, W. B. Wang, and J. H. Yong, "A video summarization tool using two-level redundancy detection for personal video recorders," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 521-526, 2008.

[28] N. Sethi and A. Aggarwal, "Robust face detection and tracking using pyramidal Lucas Kanade tracker algorithm," *International Journal of Computer Technology and Applications*, vol. 2, no. 5, pp. 1432-1438, 2002.

[29] C. Harris and M. Stephens, "A combined corner and edge detector," *Alvey Vision Conference*, vol. 15, pp. 50, 1988.

[30] B. Han, Y. Hu, G. Wang, W. Wu, and T. Yoshigahara, "Enhanced sports video shot boundary detection based on middle level features and a unified model," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 3, pp. 1168-1176, 2007.

[31] F. Zheng, S. Li, H. Li, and J. Feng, "Weighted block matching-based anchor shot detection with dynamic background," *International Conference on Image Analysis and Recognition*, pp. 220-228, 2009.

[32] C. R. Huang, H. P. Lee, and C. S. Chen, "Shot change detection via local keypoint matching," *IEEE Transactions on Multimedia*, vol. 10, no. 6, pp. 1097-1108, 2008.

BIOGRAPHIES

Carlos Cuevas received the Ingeniero de Telecomunicación degree (integrated BSc-MSc accredited by ABET) in 2006 and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications), in 2011, both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain.

Since 2006 he has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM. His research interests include signal and image processing, computer vision, pattern recognition and automatic target recognition.

Narciso García received the Ingeniero de Telecomunicación degree (five years engineering program) in 1976 (Spanish National Graduation Award) and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications) in 1983 (Doctoral Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain.

Since 1977 he is a member of the faculty of the UPM where he is currently a Professor of Signal Theory and Communications. He leads the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM. He has been actively involved in Spanish and European research projects, serving also as evaluator, reviewer, auditor, and observer of several research and development programmes of the European Union. He was a co-writer of the EBU proposal, base of the ITU standard for digital transmission of TV at 34-45 Mb/s (ITU-T J.81). His professional and research interests are in the areas of digital image and video compression and of computer vision.